Forecasting Research Division
Technical Report No.72

Quality control of aircraft data: an intercomparison, case studies and improvements

by
N Bruce Ingleby

November 1993

Quality control of aircraft data: an intercomparison, case studies and improvements

N Bruce Ingleby

Abstract

Aircraft quality control (qc) decisions from three operational centres are compared for two cases in August 1992. The interaction of manual and automatic quality control at the Met Office is investigated. Three separate examples from a case in September 1992 are examined. They are fairly typical of the problems that the buddy check has in dealing with large wind shears in data dense areas. Improvements to the buddy check, including reducing the number of buddies, are tested. An experimental whole flight check was set up, and more realistic observation and background error estimates have been implemented.

1. Intercomparison of aircraft quality control with NMC and ECMWF

Listings of flagged aircraft observations were exchanged between the UK Meteorological Office (UKMO), the National Meteorological Center, Washington (NMC) and the European Centre for Medium-Range Weather Forecasting (ECMWF). The two cases (00Z 5 August 1992 and 12Z 6 August 1992) were chosen in advance at random. The late cut-off runs were used, so that almost all data would be available at the three centres.

The numbers of observations rejected are given in table 1. NMC rejects the highest number of observations, and ECMWF the lowest, with UKMO in between. The UKMO quality control was rerun with no forecaster intervention (NI), this provides a cleaner comparison as NMC and ECMWF have little or no human intervention. The no intervention run rejects more observations than the operational run with intervention, but still rather fewer than NMC. At UKMO the totals include calm winds which are automatically flagged. The numbers of 'calms' are given, as are the total numbers of 'usable' observations for each run (available at UKMO).

	Wind		Temperature	
	0500	0612	0500	0612
ECMWF	21	18		
NMC	36	51	28	29
UKMO	19	19	15	10
(Calms	6	8)		
UKMO(NI)	32	37	18	17
Number	1220	1387	1231	1416

Table 1. Numbers of AIREP observations flagged by the different systems (PIREP and AMDAR not included).

It was noticed that the three centres had slightly different interpretations of the flight levels. These are given in table 2. The differences are less than 1 hPa and increase slightly with height - presumably due to minor differences in the constants used. For practical purposes the differences are insignificant, but along with slightly different positions (due to different beacon dictionaries, and whether positions are stored to one or two decimal places) they complicate matching of observations.

flight level feet	UKMO	ECMWF	NMC
39000	196.8	197.2	197
37000	216.6	217.1	216
35000	238.4	239.0	238
33000	262.0	262.6	262
31000	287.4	288.1	287
29000	314.8	315.5	315
27000	344.3	344.6	344
22000	427.9	428.5	428
19000	485.5	485.8	485

Table 2. Flight level pressures (hPa) at the different centres (as listed).

Subjective comparison

The first finding was how few observations were in common, particularly

between the UKMO and ECMWF lists. The ECMWF lists together with comparable information from the UKMO QC are given as table 3. The main reasons for the observations being flagged by ECMWF and not by UKMO are

a) manual intervention at UKMO correcting bad observations (9 observations)

b) observations that had missing wind at UKMO (due to more sophisticated decoding programs at ECMWF?) (6 observations)

c) permanent rejects at UKMO (not usually listed) (7 observations)

The three remaining 'real' differences in quality control were due to large differences in the background wind, in one instance deriving from a difference in position. In total ECMWF only rejected four observations South of 30°N. This is too few by comparison with UKMO and NMC, and may be because the ECMWF estimated background errors are too large, being based largely on observation density (through the OI estimates of analysis error). In UKMO and NMC systems there are relatively few rejections in the tropics - this partly reflects reduced data density there as shown in figure 1. Data coverage is good over the North Atlantic, USA and Australia, and moderate over the North and Central Pacific (with some diurnal variation). Elsewhere there are huge data voids. The reports over Australia are mainly from automated AMDAR systems which are generally of very good quality (both in terms of root mean square error and percentage of gross errors).

Erroneous decisions fall into two categories: 'false alarms' where a good observation has been rejected, and 'misses' where a bad observation was not detected. (In the absence of anything better subjective evaluation will be used to decide whether observations are 'good' or 'bad'.) 'False alarms' can be detected by comparing with the background and nearby observations without reference to any other system. 'Misses' are more difficult to detect because it is impractical to subjectively examine the 95-99% of observations that pass the quality control checks. Within the Met Office system observations with a Probability of Gross Error (PGE) over 0.3 after either the background or buddy check are listed in an attempt to draw attention to 'suspect' observations. This has not been particularly successful partly because not all 'bad' observations exceed this criterion. Also very few of the listings have been carefully examined and an interactive graphical facility highlighting flagged and suspect observations would be much more effective than a listing.

An obvious use of the intercomparison was to look for erroneous decisions particularly 'misses' in the UKMO system. To some extent 'false alarms' in the other systems can be detected, but without a list of all observations from

a) ANALYSIS DATE: 05/08/92 ANALYSIS TIME: 00Z

					<------ ECMWF ------>			<------- UKMO ------->		
	Lat	Long	time	press	ob	bk		ob	bk	
FDX122	80.28	-115.00	10	288.1	270/ 60	345/ 31	FG	270/ 60	357/ 36	99F
DAL46	53.00	-50.00	118	239.0	286/ 95	283/ 55	FIN	286/ 95	283/ 52	+00
DAL143	50.68	-127.36	-42	217.1	175/ 49	90/ 4	FG	175/ 48	290/ 14	99F
FDX011	49.70	159.31	-140	217.1	240/ 70	343/ 77	FG	240/ 70	336/ 59	99F
DLH570	46.15	-122.91	-137	262.6	335/ 16	258/ 52	FG	no data		
REACH1	45.00	-50.00	110	262.6	296/ 17	291/ 82	FG	296/ 98	291/ 73	c00
XX999	43.65	-96.78	-51	288.1	245/ 72	285/ 50	FG	245/ 72	298/ 51	p9*F
DL1210	43.10	-88.26	101	239.0	319/ 60	274/ 65	FG	319/ 61	292/ 71	p0*F
DL466	42.46	-98.68	11	217.1	260/ 66	291/ 79	FIN	260/ 67	296/ 70	p7*F
UA19	42.46	-98.68	79	197.2	264/107	299/ 74	FG	no data		
UA19	42.61	-94.30	47	197.2	289/115	297/ 68	FG	no data		
UA108	42.05	-83.45	-32	262.6	330/ 53	303/ 96	FG	no data		
XX999	42.05	-83.45	120	262.6	316/ 91	291/ 96	FIN	316/ 91	299/ 60	p3*F
XX999	40.58	-88.16	19	942.1	282/ 54	36/ 11	FG	282/ 54	45/ 9	p9*F
DL626	39.11	-94.60	-84	262.6	338/ 31	285/ 56	FG	338/ 31	290/ 54	p4*F
IBE640	37.00	-40.00	-122	288.1	350/115	318/ 21	FG	350/114	297/ 22	-9*F
XX999	34.68	-92.18	-139	344.6	212/ 45	263/ 29	FIN	212/ 44	274/ 32	p5*F
CPA501	34.11	139.50	162	197.2	305/ 76	308/ 20	FG	no wind		
CPA565	27.06	124.00	176	197.2	360/ 64	316/ 13	FG	no wind		
XXX	15.00	32.00	177	376.3	250/ 39	86/ 20	DIR	no data		
JAL776	-23.00	150.00	38	197.2	0/ 0	254/ 85	FG	262/100	250/ 82	c00

b) ANALYSIS DATE: 06/08/92 ANALYSIS TIME: 12Z

					<------ ECMWF ------>			<------- UKMO ------->		
	Lat	Long	time	press	ob	bk		ob	bk	
LOT500	69.00	-10.00	50	262.6	30/ 29	220/ 25	FG	230/ 30	230/ 20	c00
VVRD23	66.00	-30.00	51	428.5	328/ 31	313/ 80	FG	328/ 31	302/ 43	00
DAE680	62.00	-40.00	-24	485.8	190/ 19	291/ 33	FG	290/ 20	313/ 25	c00
SAS911	59.00	-40.00	152	217.1	198/ 33	296/ 43	FG	298/ 33	299/ 38	c00
NWA49	58.00	-40.00	110	239.0	350/ 60	300/ 42	FG	305/ 60	302/ 41	c00
DAL63	57.00	-20.00	67	239.0	235/ 52	266/ 14	FG	235/ 53	229/ 22	+07F
DAL52	50.00	-160.00	133	217.1	290/ 29	287/115	FG	290/ 30	286/119	-9*F
IBE400	46.00	-30.00	130	217.1	230/ 91	313/ 93	FG	320/ 91	315/ 78	c00
IBE640	46.00	-30.00	130	217.1	230/ 91	313/ 93	FG	320/ 91	315/ 78	c00
JAL12	45.13	16.53	3	239.0	350/ 80	7/ 33	FG	350/ 80	13/ 33	-9*F
DLH546	45.00	-20.00	134	288.1	200/ 29	296/ 42	FG	200/ 30	252/ 48	-0*F
CCA991	44.00	160.00	44	239.0	240/ 60	334/ 69	FG	340/ 60	340/ 70	c00
CAL006	42.00	170.00	128	239.0	125/ 70	202/ 52	FG	125/ 70	218/ 47	99F
FDX191	41.17	5.01	49	315.5	330/ 41	114/ 12	FG	different position		
FDX191	41.00	-30.00	49	314.8				330/ 40	320/ 46	00
KAL016	40.63	165.00	81	288.1	10/ 45	144/ 26	FG	10/ 45	27/ 17	00
CCA991	39.58	150.53	-16	239.0	230/ 29	309/ 32	FIN	320/ 30	311/ 32	c10
JAL62	39.00	160.00	-54	262.6	260/ 76	9/ 78	FG	360/ 75	6/ 64	c00
LCO391	-29.50	-61.55	-75	197.2	275/200	275/ 63	FG	no data		

Table 3. Aircraft reports rejected by ECMWF: callsign, latitude, longitude, observation-analysis time (minutes), pressure (hPa), observation and background wind (degrees/knots), ECMWF flags - First Guess, FINal or DIRection. The UKMO quality control information consists of a character (+ supported, - rejected, c corrected) the probability of gross error (PGE) after the background check, the PGE after the buddy check, followed by an F if the wind has a final flag set. The PGEs are multiplied by 10; 0-4 pass the qc, 5-9 and * fail.

the other centres it is not possible to say whether their quality control 'missed' a particular bad observation or if it was absent for a different reason.

As mentioned above comparison with the ECMWF list did not reveal any 'misses' in the UKMO quality control. Comparison with the NMC list was more interesting. The NMC list only gave observation callsigns and positions, so no comparison of background values is possible. I had the impression that UKMO temperature checks were too lax. In some cases the NMC checks particularly the superob check rejected winds which had directions 10° different to a consensus of the other winds present. AIREPs often report wind directions to the nearest 5 or 10 degrees, and in strong wind cases this leads to significant truncation errors. For the first case 6 of the wind observations failed by NMC but passed in the UKMO system looked wrong or highly suspect and 8 appeared probably correct. These numbers should be taken as approximate because the situation is considerably complicated by the human intervention at UKMO - correcting, supporting and rejecting some of the observations on the NMC list.

One example of a dubious wind direction, in a data sparse area, which was detected by the NMC quality control but which passed the UKMO (and presumably the ECMWF) system is that from NCA082 in the centre of figure 2. It is plotted against height contours and isotachs (dotted) from the forecast background at 250 hPa. In this and following figures dotted wind shafts indicate winds that failed the quality control, dashed shafts indicate suspect winds (marginal pass) and solid shafts the winds which easily passed the quality control.

2. Intervention aspects

The forecaster intervention at UKMO has already been mentioned. Forecasters have a facility to support elements of an aircraft report - the observations are still subject to the automatic quality control, but with a reduced initial probability of gross error, and the estimate of the background error is increased (only in the background check for the observations considered here - on 1 December 1992 this was extended to the buddy check and analysis). In the two intercomparison cases four supported observations were rejected by the automatic quality control. These were investigated as were the three from the 12Z 28 September case (next section). They are listed in table 4, and commented on below.

	lat	long	time	P(hPa)	T(°C)	o-b		DD/FF(kts)	b	
00Z 5 Aug 1992										
UA133	43.0	-83.7	117	238.4	52.0	103.5	**F	309/ 95	299/ 51	+56F
UAL0	38.7	-81.9	5	216.6	-56.0	-2.3	34	87/ 98	270/ 79	+**F
KAL851	24.0	-160.8	-48	287.4	-30.0	5.1	11	220/ 30	330/ 36	+08F
12Z 6 Aug 1992										
DAL63	57.0	-20.0	67	238.4	-55.0	-3.5	00	235/ 53	229/ 22	+07F
12Z 28 Sept 1992										
CDN4	37.0	144.0	-157	216.6	-53.0	-0.5	21	203/136	288/ 99	+99F
KAL072	38.3	147.2	62	216.6	-52.0	0.8	00	300/136	289/ 98	+07F
KAL052	35.5	150.0	65	216.6	-53.0	0.4	00	310/105	282/ 77	+06F

Table 4. Supported observations flagged by the automatic quality control. See table 3 for explanation of listing, for temperatures the observed and observation-background values are given.

UA133 was at the edge of a jet and the wind is probably correct, however the sign error in the temperature combined with the large wind increment caused the observation to be given a probability of gross error in location of over 0.5 and hence both temperature and wind were rejected. At this analysis time there were 7 temperature sign errors all from United Airways aircraft.

UAL0: as it stands this observed wind is about 180° in error. On investigation it was found that the forecaster had tried to apply both a correction and a supporting flag, but the interface was not set up for this combination and the correction was not implemented. It was decided that supporting corrected observations could be dangerous and the software has since been changed to ignore the support and process the correction. In this case there is some doubt as to whether the correct wind was from 267° (a 180° error) or 287° (the attempted correction).

KAL851: the background is approximately correct and the observation should presumably have been 330° rather than 220°.

DAL63: again the background is approximately correct and it appears that the forecaster supported the observation without checking both speed and direction.

CDN4: the observation has a wind direction error, probably of 100°.

KAL072: the observation was probably correct, but it disagreed with earlier buddies as the wind strengthened and veered in the meantime (see section 3.3).

KAL052: the observation was probably correct (see section 3.3).

One rather curious correction was also found at 12Z on 6 August. UAL828 at 45°N and 170°E had a reported wind of 278/61, the background was 232/59 and the observation was changed to 330/95.

At 12Z on 28 September four surface BOGUS observations located around 52.7°N and 81.7°E had differences from background of 35 hPa, and were obviously mispositioned.

The man-machine mix

Because of the vast amount of data available every six hours an entirely manual approach is out of the question, but an almost entirely automatic approach is feasible (as at ECMWF). For economic reasons the level of manual quality control is likely to decline rather than increase. It is thus essential that the automatic quality control should be as comprehensive and reliable as possible, and that we use the manual effort to its best effect, including providing feedback on the automatic qc. **Manual quality control should build on the automatic quality control, not be independent of it.**

Unfortunately the current manual quality control is largely separate from the automatic qc. Before the analysis the intervention forecasters only see 'flags' from a rather simple background check, rather than from the operational quality control, as a result they intervene in cases that the operational automatic qc is quite capable of handling. Ingleby (1991) found that for aircraft winds and temperatures over half of the reports rejected by CFO would have been rejected by the background check anyway. It is only possible to assess the situation for supported observations by rerunning cases without the intervention applied. This has been done for six global cases by S Waters (1992, unpublished report on 'Intervention on aircraft data'). He found that 'A large amount of the current intervention effort is "wasted" supporting observations which will not be flagged by quality control, and otherwise duplicating the automatic system.'

A much more integrated approach has been proposed, in which 'preliminary runs' of the automatic qc are performed half an hour before the final cut-off. The system is now running in trial mode once a day. The observations are displayed colour coded and the forecasters can concentrate on those observations which are either flagged by the automatic qc or borderline passes. Unfortunately forecaster use of the preliminary run has not been great, because of unfamiliarity, satisfaction with the current system and the fact that much of the intervention has been performed by the time that the preliminary run is available. At the final cut-off time the data pass through the automatic qc again together with any observations received since the preliminary run (the extra data may occasionally result in differences from the preliminary QC buddy check decisions). Automatic quality control of BOGUS and supported observations is necessary as some of the examples above make clear.

3. Buddy check problems

The 12Z run on 28 September contains three good examples of problems with the buddy check of aircraft winds where there is strong wind shear in a data dense area. These are examined in this section. It is worth pointing out that there were no such problems apparent in the two intercomparison cases examined earlier.

3.1 Example 1. UAL852: 47°N, 160°W, 12Z 28 September 1992

D Ireland (CFO) pointed out that UAL852 had been incorrectly rejected by the buddy check and suggested that it was due to the presence of an erroneous wind from CCA981 nearby. A detailed examination of the buddy check for this case is included as it illustrates several aspects of the buddy check.

Nearby observations are listed in table 5, those which were checked against UAL852 have an estimated correlation, an agreement factor and a multiplication factor shown. The estimated background error correlation is a decreasing function of the horizontal, vertical (log pressure) and time separation between the two observations. The agreement factor depends on the correlation, the estimated observation and background accuracies and the observation increments (differences from background); it is independent of the estimated probabilities of gross error (PGEs). The multiplication factor is

QG12 RUN 12Z ON 28/ 9/92 AIREPS

callsign	lat	long	time	P(hPa)	DD/FF o	DD/FF b		corr	agree	mult
UAL810	47.1	-160.2	112	216.6	316/164	327/140	10			
NWA26	47.0	-160.0	-48	196.8	320/145	324/128	+00			
UAL808	47.0	-160.0	-9	196.8	320/170	324/128	+00			
UAL880	47.0	-160.0	18	216.6	320/160	327/137	+00			
UAL828	47.0	-160.0	32	216.6	315/158	327/137	+00			
DAL78	47.0	-160.0	72	216.6	316/169	327/137	+00			
UA806	47.0	-160.0	96	216.6	320/160	327/137	+00	.944	0.125	2.81
UAL806	47.0	-160.0	96	216.6	320/160	327/137	+00	.944	0.125	2.77
UA810	47.0	-160.0	112	216.6	315/160	327/137	+00	.946	0.041	4.20
UAL810	47.0	-160.0	112	216.6	315/160	327/137	+00	.946	0.041	3.06
JAL2	47.0	-160.0	132	216.6	320/160	327/137	00	.944	0.124	1.80
CCA981	47.0	-160.0	12	238.4	305/147	329/147	99F	.921	0.003	1.01
NWA24	47.0	-160.0	28	238.4	325/140	329/147	03	.943	1.284	0.93
DAL50	47.0	-160.0	91	238.4	326/145	329/147	00	.996	1.279	0.92
UAL852	47.0	-160.0	115	238.4	330/140	329/147	05F			
DAL74	47.0	-160.0	132	238.4	325/140	329/147	00	.998	1.310	0.91
NWA2	47.0	-160.0	10	262.0	335/160	330/151	+01	.871	0.834	1.06
DAL52	47.0	-160.0	87	262.0	335/174	330/151	+04	.942	0.213	1.41
AAL128	47.0	-160.0	107	262.0	335/170	330/151	+01	.948	0.325	1.26
NWA28	45.0	-160.0	50	196.8	310/165	316/140	+00	.574	0.982	1.00
KAL002	45.0	-160.0	-55	216.6	300/140	320/150	+00	.560	1.055	0.99
UAL844	45.0	-160.0	23	216.6	310/160	320/150	+00	.656	1.078	0.98
UA890	45.0	-160.0	47	216.6	310/150	320/150	00	.676	1.095	0.98
UAL890	45.0	-160.0	47	216.6	310/150	320/150	00			
KAL018	45.0	-160.0	30	238.4	320/140	324/161	00			
JAL62	45.0	-160.0	95	238.4	320/135	324/161	00			
ANA006	45.0	-160.0	111	238.4	320/145	324/161	00			
MAS92	45.0	-160.0	175	238.4	320/150	324/161	00			
UAL844	44.7	-158.1	31	216.6	313/158	315/137	00	.558	0.874	1.03

Table 5. Listing of observations at and close to (47N, 160W).
Most of the columns are as for table 3 (but no ECMWF data).
'corr', 'agree' and 'mult' give the background error correlation,
the agreement factor and the multiplication factor for observations
buddy checked with UAL852. See text for explanation.

the factor by which both PGEs are multiplied to update them; it is a function of the agreement factor and the current PGEs. The agreement factor is greater than 1 if the two observations support each other, less than 1 if the two observations disagree, *vice versa* for the multiplication factor. They are defined by equation 4.15 and the reciprocal of 4.16 respectively of Ingleby and Parrett (1993). (4.16 is raised to the power 0.5 by the buddy check damping - this is included in the multiplication factors shown.) Note that UAL806 is a duplicate of UA806 and has the same correlation and agreement factor, but a smaller multiplication factor (as a result of the PGEs being updated), similarly UAL810 is a duplicate of UA810.

The buddy check was performed sequentially in the order shown in table 5, ie UA806 was the first to buddy check UAL852, and UAL844 the last. It can be seen that although CCA981 had a very poor agreement with UAL852 (0.003), the high PGE of CCA981 prevented it having much effect - the multiplication factor was only 1.013. Much more significant were the observations at 216.6 hPa which gave large multiplication factors. The other observations at 238.4 hPa (except CCA981) supported UAL852 but not strongly. Lower level observations at 262.0 hPa tended to disagree, while those two degrees further south had smaller correlations and little effect. NWA24 at the same level and DAL52 below also had their PGEs increased by the buddy check until they were considered suspect.

Figures 3a) and 3b) show the observations and their differences from background respectively as time-height plots. At each level the observations are fairly consistent, but there is a double jet core structure with the observations at 238.4 hPa about 20 knots weaker than those above and below (figure 3a). The observations at 238.4 hPa (except CCA981) are quite close to the background whereas those above and below have larger increments (figure 3b). This is an example of a known feature of the pairwise buddy check that observations with small increments are given insufficient weight relative to reports with larger increments. This behaviour can be reduced by reducing the number of buddies - see section 4.1.

The climatologically estimated background error correlations between UAL852 and the observations above and below are very large, ~ 0.945 (table 5), indicating that on average we expect substantial consistency between the background errors at the different levels. Figure 3b) suggests that the actual background error correlation between 216 hPa and 238 hPa is very small, and that the correlation between 238 hPa and 262 hPa is actually negative. An

experiment in which the background error correlations were reduced is reported in section 4.2.

The question of how much the model can support the resolution in the observations arises. The model has layer mid-points at approximately 300, 250 and 200 hPa. Thus the 262.0 and 238.4 hPa aircraft observations are within the same model layer, and the observations at 216.6 hPa fall just outside it, and it becomes clear that the double jet structure cannot be resolved.

3.2 **Example 2.** AMDARs: 38°S, 145°E, 12Z 28 September 1992

While examining the 12Z 28 September case it was noticed that 20 AMDAR observations (automated aircraft reports) had been rejected by the buddy check. 19 of these were from three different flights at low levels (take-off or landing) over Melbourne (~ -37.7, 144.8). The time height plot of observed winds (figure 4a) shows considerable backing during the six-hour analysis window. This is confirmed by the 850 hPa height field from the background forecast (figure 5) which shows a sharp trough with its axis over Melbourne at 12Z. This consistent picture and the usually excellent quality of AMDAR data suggest that all the observations are correct. However it is not surprising that the buddy check had problems with the opposing observation increments that were present in the first hour (figure 4b).

Here the standard background error correlation in time is completely inappropriate whereas there is good consistency in the vertical - the opposite of the first example. In this example interpolation of the background fields in time would have helped, as the background seems to have a good representation of the situation. However the situation would not be properly resolved by the analysis which inserts data over a five hour window.

3.3 **Example 3.** KAL072 (38.3, 147.2), KAL052 (35.5, 150.0), 12Z 28 Sept 1992

This is another jet stream case involving rejections by the buddy check, this time East of Japan. Both these observations were supported by CFO but rejected by the buddy check. KAL072 is just on the southern flank of a jet core, while KAL052 is south-east of the jet core (figure 6, note that only observations from the last three hours of the analysis window are plotted). KAL072 does not agree well with other observations at the same point (figure 7) and is much stronger than the two earlier reports at the same level. However it is possible that the jet core has moved over this position, and the

reported wind of 300/136 is reasonably consistent with the observations 1.5 degrees upstream. The background wind at the level of KAL072 is 289/98 which is 7 and 10 knots weaker than that at the two flight levels below. All this suggests that the situation can only be modelled satisfactorily by a four dimensional representation, and that it is unsurprising that the (essentially three dimensional) quality control has difficulties.

KAL052 is at a higher flight level than all the other observations at the same point, and is rather stronger than the others (figure 8).

4. Quality control tests/changes

4.1 Reduction in number of buddies

With an ideal quality control system reducing the number of buddies reduces the amount of information available and would be detrimental. The previous OI based quality control described by Bell and Dickinson (1987) only used up to seven buddies, and one of the advantages of the Bayesian quality control with a pairwise buddy check introduced by Lorenc and Hammon (1988) was that it was computationally cheaper and so more buddies could be used. When the Bayesian scheme was first implemented operationally for surface marine observations in March 1988 it was found necessary to reduce the maximum number of buddies from 20 to 12 (section 5.1 of Parrett, 1988) but no theoretical explanation was available at the time.

In the pre-operational testing on the Cray computer in the first half of 1991 some rejection of good observations close to the background (and occasional passing of poor observations further from the background) in data dense areas was found. Through the use of simple examples and an examination of the theory it became clear that the approximation of just comparing two observations at a time was largely responsible, and that as the number of buddies increases the approximation becomes worse. Buddy check 'damping' was introduced to curb over-activity of the buddy check (see section 2.2 of Ingleby (1991) for a brief description of cases, and section 4.6 of Ingleby and Parrett (1993) for the equations used). 'Damping' and reducing the number of buddies can reduce but not eliminate the buddy check problems in data dense areas. In the medium term it is desirable to move to an alternative method of quality control which does not share this inherent weakness. Ingleby and Lorenc (1993) made a theoretical study of methods of quality control, which clarifies the approximations used in practice.

The method by which buddies are chosen is described in Ingleby and Parrett (1993). Basically the observations are sorted by 5° latitude band, longitude, latitude, pressure and time. An observation could be selected by observations preceding it in the list and could itself then select up to N_{bud} subsequent buddies including up to N_{band} from the same band. There is no set limit on the number of preceding observations that can select it, but an observation in a data dense area will typically have $2 \times N_{band}$ close buddies (buddies from other bands usually have less effect). At the time of this case N_{bud} = 12 and N_{band} = 8, table 6 shows the effect of reducing these parameters, the last column (0/0) is the background check.

N_{bud}	16	14	13	12	11	9	8	6	5	4	0	
N_{band}	11	10	9	8	7	6	5	4	3	2	0	
AIREP Temperature	12	11	10	9	9	9	9	7	8	7	11	
AIREP Wind	21	19	18	17	13	13	13	13	11	10	11	
AMDAR Wind	22	22	22	20	18	14	12	8	4	4	0	
NWA24	6F	5F	4	3	3	1	0	0	0	0	0	
UAL852	8F	8F	7F	5F	4	2	0	0	0	0	0	
DAL52	7F	7F	6F	4	1	1	0	0	0	0	0	+
KAL072	6F	7F	7F	7F	7F	7F	6F	6F	5F	6F	0	+
KAL052	8F	8F	8F	6F	4	2	5F	7F	6F	4	0	+

Table 6. Numbers of observations rejected (top) and selected individual PGEs (bottom) as the numbers of buddies are decreased (see text for details). The first three observations are at (47.0, -160.0) - see section 3.1, KAL072 and KAL052 are at (38.3, 147.2) and (35.5, 150.0) respectively - see section 3.3. The last three observations (marked +) were supported by CFO. The boxed columns refer to the then operational and the revised parameters.

Although the reduction in number rejected as N_{band} is reduced looks consistent (except at very low numbers of buddies) this hides some changes in the set of observations rejected, with some observations, eg KAL052 in the table, changing from fail to pass and then back to fail as the number of buddies is reduced. This illustrates one of the weaknesses of the buddy check

- an apparent arbitrariness dependent on the exact selection of buddies used.

Following this investigation the parameters were reduced to N_{bud} = 9 and N_{band} = 6 with effect from 1 December 1992. This is a fairly modest reduction, but is seen to give an worthwhile improvement in this case: UAL852 and KAL052 pass the buddy check, and the number of AMDARs rejected reduces from 20 to 14. Only KAL072 is unaffected by changing the number of buddies, reference to figure 7 shows that all its geographically closest buddies disagree with it, as mentioned in section 3.3 observations upstream give it more support. Increasing the number of buddies was tested for comparison (see table 6), and does give worse results, failing two extra AMDARs and two more aircraft from the first example (section 3.1).

The reduction in the number of buddies was also tested on surface observations, where it had a more mixed effect - improving some decisions and making others worse - with a narrow majority for the improvements. As expected the test decisions were closer to background decisions in general, and all occurred in data dense areas. The reduction in the number of buddies was applied to all observations that are buddy checked: aircraft, surface, radiosondes and satellite cloud track winds (radiosondes and 'SATOBs' do not occur in such dense clusters and so the effect on them is minimal).

4.2 Modified vertical correlations

In order to try to model better the narrow vertical error correlations associated with jet streams an experiment was carried out using a vertical correlation based on potential temperature (θ in Kelvin) differences, rather than on log(pressure) differences. The function tested was

$$\mu = 0.32 + 0.68(1+0.19|\Delta\theta|)\exp(-0.19|\Delta\theta|)$$

as given in equation (4) of Benjamin (1989). The use of a θ based correlation was suggested by A Lorenc and is more physically based than a statistically tuned dependence on wind speed or wind shear (see comments on isentropic analyses in the introduction of Benjamin (1989)). This function was used for test purposes only, and would need to be validated/tuned before being adopted in our operational system. In particular the constant term (0.32) is very large and unsuitable for observations widely separated in the vertical, where the correlation usually tends to zero.

This function did indeed narrow the correlations in jet situations. In the first example (section 3.1) the correlation between UAL852 and the

observations above it dropped from ~ 0.945 to ~ 0.740. As a result the number of rejections reduced slightly and the PGEs of UAL852 and nearby observations were decreased (compare table 7 with table 6). This is in broad agreement with the significant reduction in flagging of surface pressure observations when correlations were artificially reduced (table 1 of Ingleby, 1991). Note however that the number of AMDARs rejected increased by 1 for some of the combinations of N_{bud}/N_{band}.

N_{bud}	12	11	9	8
N_{band}	8	7	6	5
AIREP Temperature	9	9	9	9
AIREP Wind	14	12	12	12
AMDAR Wind	20	19	15	12
NWA24	0	0	0	0
UAL852	1	0	0	0
DAL52	1	0	0	0
KAL072	6F	6F	5F	5F
KAL052	6F	4	2	5F

Table 7. As table 6 except using the θ based vertical correlation described in the text.

4.3 Modified background errors

For CFO supported observations the initial PGE is multiplied by 0.2 and the background error variance is multiplied by a factor h (equivalent to multiplying the standard deviation by √h) within the background check only. For most of 1992 h had the value 4, on 1 December 1992 a change was made to enhance the background error fields in the vicinity of supported and BOGUS observations, h was correspondingly reduced to 3. This is part of the package of changes described by Parrett (1992). The 12Z 28 September case was rerun with modified background error fields and h = 4, 3, 2 and 1. Results for h = 4 and 1 are shown in table 8.

N_{bud} a)	12	11	9	8	b)	12	11	9	8
N_{band}	8	7	6	5		8	7	6	5
AIREP Temperature	9	9	9	9		9	9	9	9
AIREP Wind	14	12	11	12		14	11	12	12
AMDAR Wind	20	20	15	13		20	20	15	13
NWA24	5F	4	2	0		5F	4	2	0
UAL852	5F	4	2	0		5F	4	2	0
DAL52	3	1	1	1		3	1	0	0
KAL072	7F	6F	6F	6F		8F	8F	7F	7F
KAL052	8F	6F	5F	6F		9F	7F	6F	7F

Table 8. As table 6 but using the modified background error fields, a) h = 4, b) h = 1

The effect of the modified background fields can be seen by comparing table 8a with table 6, the experiments are identical in other respects. The effect is modest with a small reduction in the number of AIREP winds flagged, and a small increase in the number of ASDAR winds flagged. Among the selected observations the effect is most noticeable on KAL052 which consistently fails the buddy check in table 8a.

Changing h from 4 to 1 gives even more modest effects (compare table 8b with 8a, intermediate values of h are not shown because of the lack of sensitivity). With N_{bud}=9 and N_{band}=6 there is one extra supported observation flagged. If the same experiment is run with unmodified (climatological) background errors then changing h from 4 to 1 flags one extra supported observation and reinstates two unsupported observations, all in the region of example 3 (section 3.3). Particularly with the unmodified background error fields reducing h caused background check PGEs over 0.5 for some supported observations but these were reduced to near zero by the buddy check.

The upper-air synoptic dependent background errors described by Parrett (1992) will be introduced operationally on 16 November 1993. This will increase the estimate of background error near jet-streams and reduce it in low wind-speed areas. For temperature the error estimates are based on temperature gradients and tendencies.

4.4 Whole flight checks

Occasionally aircraft are subject to errors that affect all wind or temperature observations from a particular flight. Examples have included data for the wrong day being transmitted, failure of the inertial navigation system resulting in a constant wind being reported (S Waters, P Julian, personal communication), and, more commonly particularly over the US, decimal point errors in temperatures resulting in eg -53°C being reported as -5°C. While the wildly wrong values will fail the quality control anyway the possibility remains that some of the errors will pass.

Ingleby and Parrett (1993, section 5.3.2) describe an experimental check for errors affecting the whole flight. A probability of gross error in either all winds or all temperatures is calculated assuming that the reports are sufficiently separated so that the background errors are uncorrelated. Testing indicates that the check is quite good at detecting 'interesting' situations, but while some of these are due to observational errors, others are caused by large and/or systematic background errors.

As an example on 18Z 28 January 1993 UAL645 had a mean temperature difference from background of -3.8°C. It would have been soundly rejected by the check and at first sight this appeared reasonable. However if the aircraft in the vicinity are examined in narrow vertical bands, UAL645 flying at 238.4 hPa agrees quite well with other flights between 275 and 225 hPa which show a generally negative bias (figure 9a). By contrast between 225 and 175 hPa the bias tends to be positive (figure 9b). It is assumed that the model has not resolved the sharpness of the tropopause; there are no radiosonde observations over North America at 18Z preventing further checking.

In figure 9 there are two flights (UAL734 and UAL506) with temperature increments of over 40°C, an example of the decimal point error mentioned above. Both flights have been overplotted because CFO corrected the values of -5 and -6 to temperatures between -50 and -56. The corrections are unnecessary because they are taking information from the background or other observations and potentially damaging if done badly.

An apparent temperature bias was also noted for the ASDAR system PH012U (now KL012U). Comparison with Forrester and Human (1993) and other results show that its cruise level temperatures (now blacklisted) have had a warm bias of about 1.7°C over many months. However it shows little bias in ascent and descent emphasizing the difficulties of a universal approach.

4.5 Reduction of observation errors

Originally the rms observation errors were specified as 4 m/s in wind components and 1.7°C in temperature for all/most tropospheric levels. Forrester and Human (1993) verified forecasts against ASDAR reports, and based on their results low level and cruise level wind errors were reduced to 2.7 and 3.8 m/s respectively (the cruise level value will be further reduced to 3.6 m/s on 16 November 1993). For automatic aircraft reports temperature errors were reduced to 1.3°C for levels above the boundary layer. Manual temperature reports are rounded or truncated to integer values, their error estimates were unchanged. These revised values (given in full in Appendix 1 of Ingleby and Parrett (1993)) were introduced operationally on 3 August 1993.

5. Display of observations

For use in this and future quality control studies an interactive observation plotting program was set up - with it the case studies for this report are still quite time consuming, but with the tools available before they would not have been attempted. It is very difficult to determine in advance which charts will be most useful, and whether a particular decision requires a quick glance or a more substantial investigation. To investigate particular decisions it is necessary to plot all or most observations in the vicinity - in data dense areas it is useful to zoom in more than for data sparse areas. Aircraft often report 'on top of' each other making interpretation very difficult with a conventional plot. In this case the ultimate zoom-in is a time-height plot, used extensively above.

The other innovative feature is plotting of observation minus background, O-B, values in addition to, not instead of, plotting observed values. O-B values are of interest because they form the input to the quality control and analysis schemes; for instance they emphasize vector wind differences rather than the speed and direction differences more familiar to forecasters. Particularly for upper-air observations, using background values interpolated to the observation locations makes O-B values easier to compare than observed values. Just consider comparing an observed aircraft temperature at 216 hPa with a forecast field at 250 hPa and observations at other levels, similar difficulties apply to winds in strong shear situations.

Summary

The aircraft quality control at ECMWF is less strict than that at the UK Met Office, which is less strict than that at NMC. Subjective evaluation suggests that the 'truth' lies somewhere between the Met Office and the NMC performance. The intercomparison was complicated by differences in reports received, the processing applied, manual intervention and the listings provided, as well as the model background fields.

Any efficient use of manual intervention must involve the forecasters improving on the automatic quality control rather than acting independently of it, as at present.

The Met Office quality control of aircraft data performs well in general. From time to time good observations are rejected in data dense areas with strong wind shear (particularly around jet streams). In at least some of these cases the structures observed cannot be resolved by the model and/or the analysis, also any effect on the analysis is small because of the density of other observations. However reducing the maximum number of buddies (from ~ 16 to ~ 12) is beneficial, and was implemented operationally on 1 December 1992. Making the vertical correlations a function of potential temperature rather than pressure shows promise, but further work would be required to validate/tune the function used.

A whole flight check has been set up, but it is difficult to distinguish automatically between observation and model problems and it has not been implemented operationally. Improvements have been made to the specification of both background and observation errors.

Acknowledgement

Thanks are due to P Unden and P Julian for supplying the listings from ECMWF and NMC respectively.

References

R.S.Bell and A.Dickinson 1987 The Meteorological Office operational numerical weather prediction system. Meteorological Office Scientific Paper No. 41. HMSO

S.G.Benjamin 1989 An isentropic mesoα-scale analysis system and its sensitivity to aircraft and surface observations. Mon Wea Rev,**117**,7,1586-1603

D.A.Forrester and J.Human 1993 Verification of NWP model wind and temperature forecasts using ASDAR reports. Forecasting Research Division Tech. Rep. 47

N.B.Ingleby 1991 Tuning and performance of the atmospheric quality control. Short-Range Forecasting Research Division Tech. Rep. 6

N.B.Ingleby and A.C.Lorenc 1993 Bayesian quality control using multivariate normal distributions. Q.J.R.Meteorol.Soc. **119**,1195-1225

N.B.Ingleby and C.A.Parrett 1993 Quality control of atmospheric data. Version 2. Unified Model Documentation Paper 32

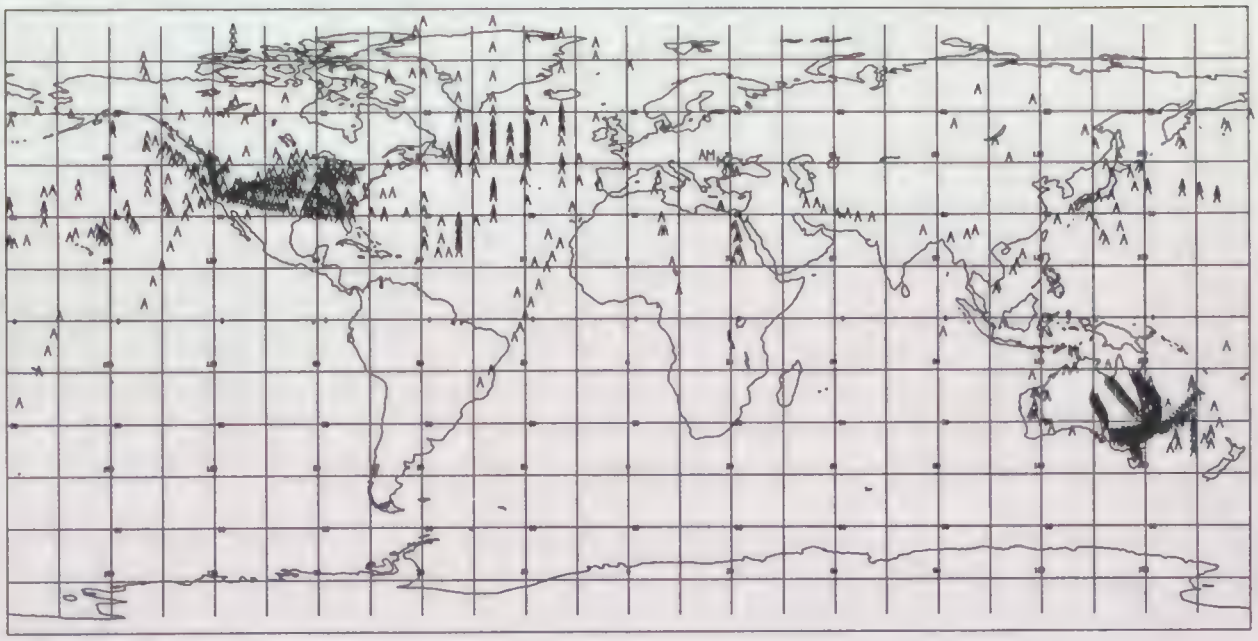
A.C.Lorenc and O.M.Hammon 1988 Objective quality control of observations using Bayesian methods. Theory and a practical implementation. Q.J.R.Meteorol.Soc,**114**,515-543

C.A.Parrett 1988 The automatic quality control of surface observations from ships. Met O 11 Tech Note 10

C.A.Parrett 1992 Background errors for the quality control and assimilation of atmospheric observations in the Unified Model - the situation in July 1992. Forecasting Research Division Tech. Rep. 22

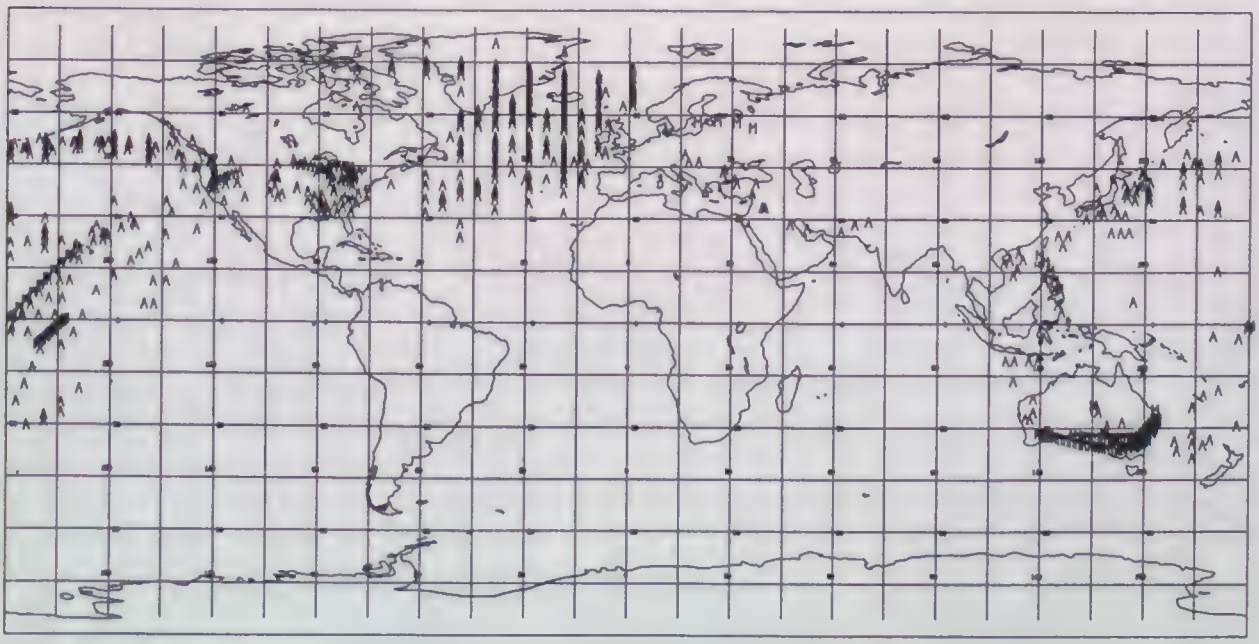
Figure 1. Distribution of aircraft reports

a) OPD OBSERVATIONS 350.0 MB TO 150.0 MB.
OB TYPE T - 180 TO T + 180
00Z 5 / 8 / 1992.



1505 OBS 1505 PLOTTED
1166 A AIREP 339 M AMDAR

b) OPD OBSERVATIONS 350.0 MB TO 150.0 MB.
OB TYPE T - 180 TO T + 180
12Z 6 / 8 / 1992.



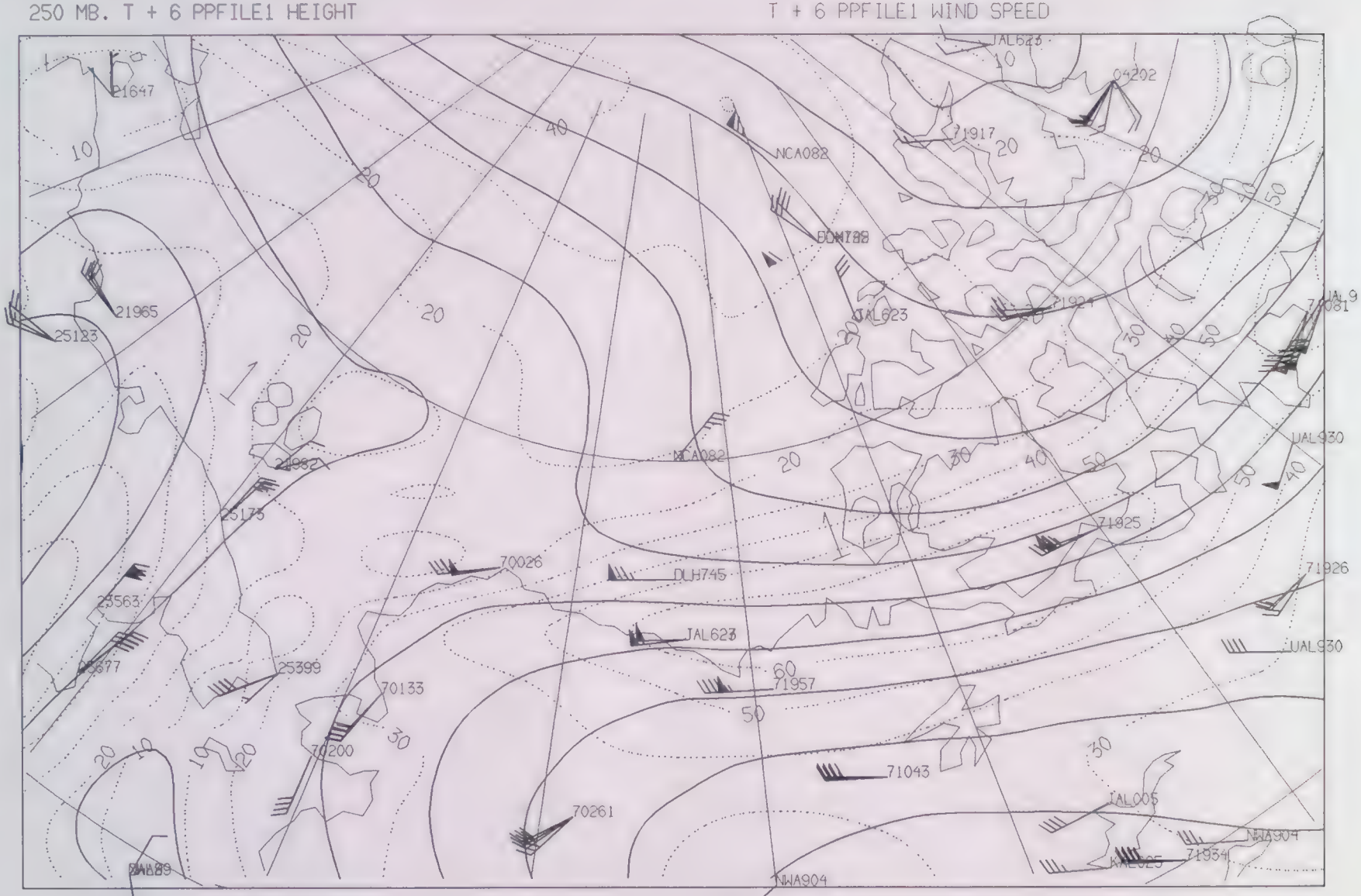
1552 OBS 1552 PLOTTED
1361 A AIREP 191 M AMDAR

OPD OBSERVATIONS 350.0 MB TO 150.0 MB. ALL
CALLSIGN 04AUG92:21.00 - 05AUG92:03.00
00Z 5 / 8 / 1992. GRID CENTRE 75.0 -141.3 SCALE: 6.00

Figure 2. Example of observation rejected by NMC and passed by UKMO. (NCA082).

250 MB. T + 6 PPFILE1 HEIGHT

T + 6 PPFILE1 WIND SPEED



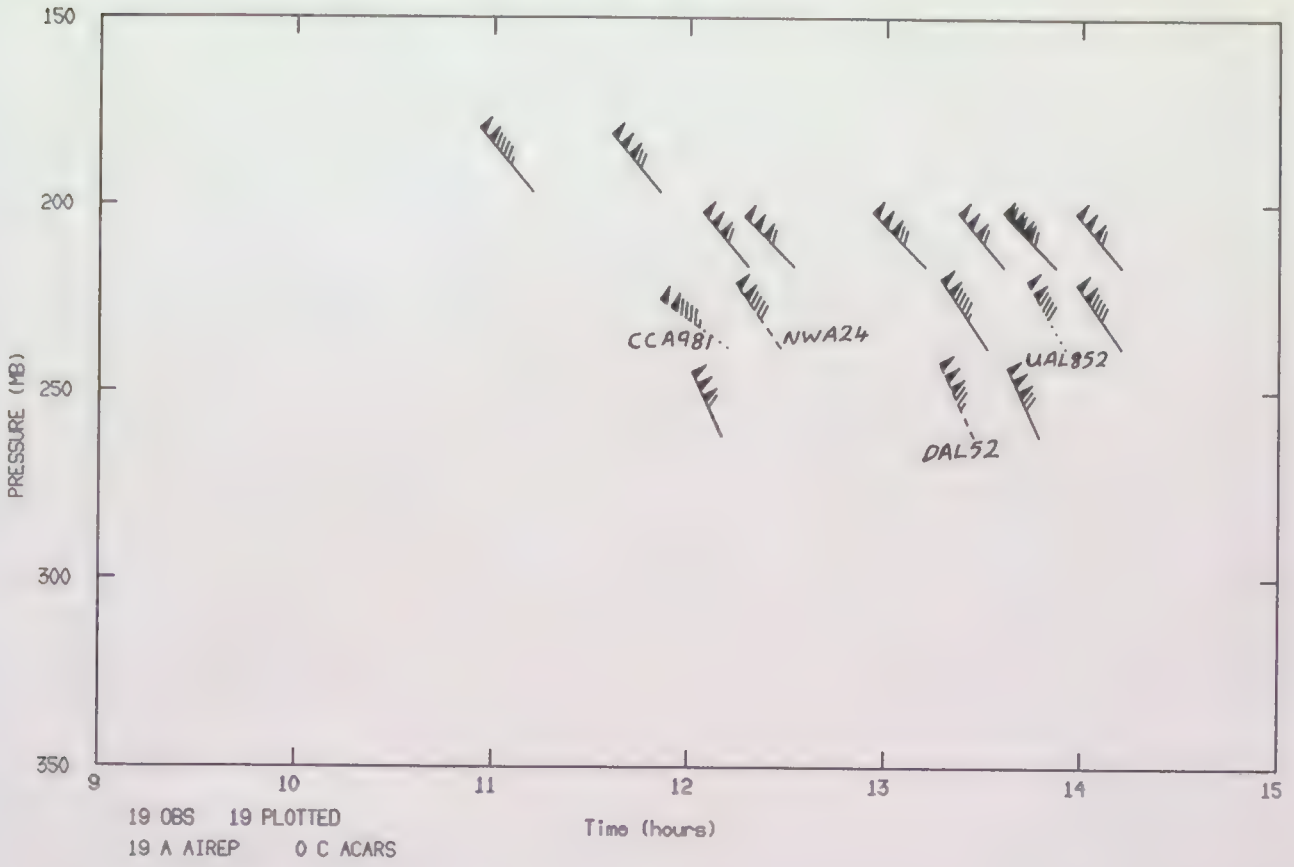
CONTOUR INTERVAL: 60 M CONTOUR INTERVAL: 10 KNOTS 110 OBS 110 PLOTTED

82 R TEMP 4 P PILOT 24 A AIREP 0 C ACARS

dotted shaft - reject, dashed - suspect

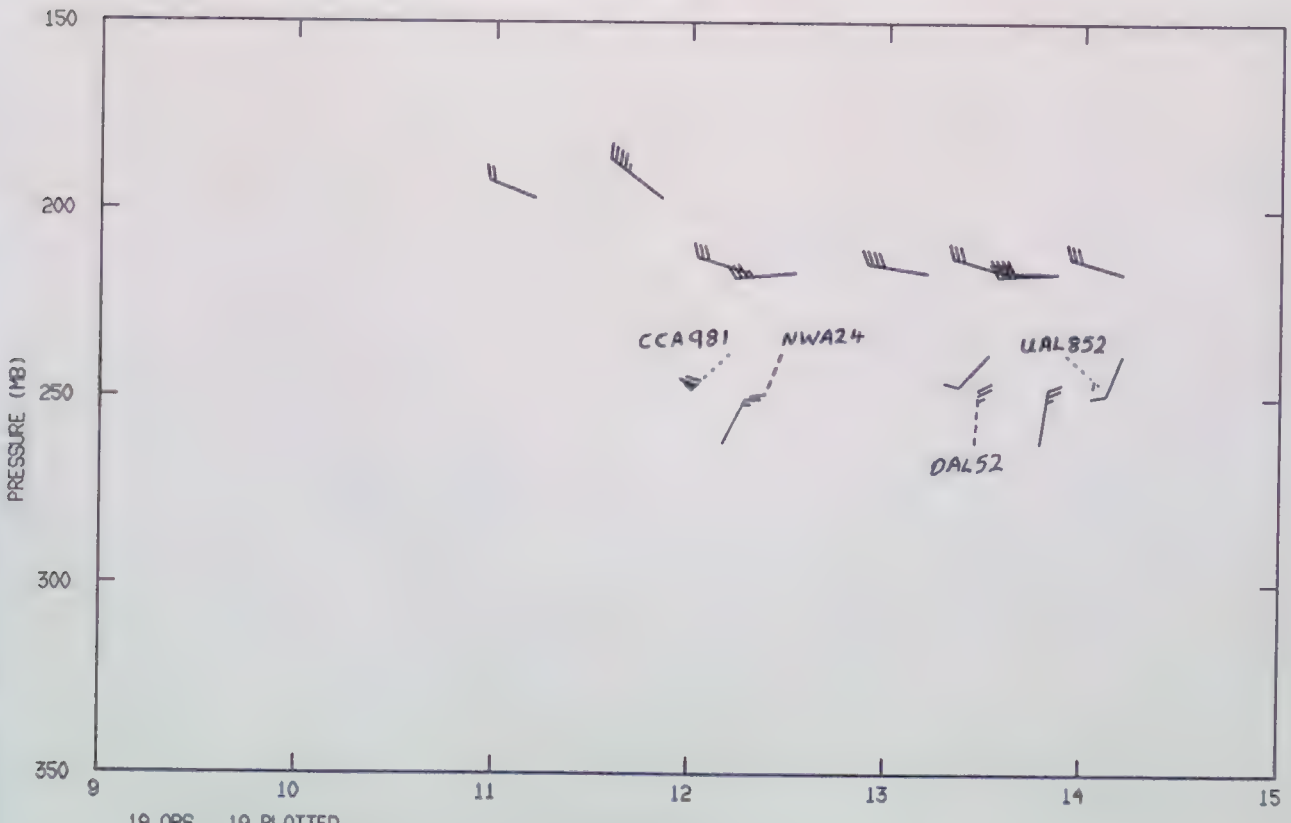
Figure 3. UAL852 and neighbouring obs.

a) OPD OBSERVATIONS 350.0 MB TO 150.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE 47.0 -160.0 AREA: 1.10



19 OBS 19 PLOTTED
19 A AIREP 0 C ACARS

b) OPD (OBS. - BK.) 350.0 MB TO 150.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE 47.0 -160.0 AREA: 1.10

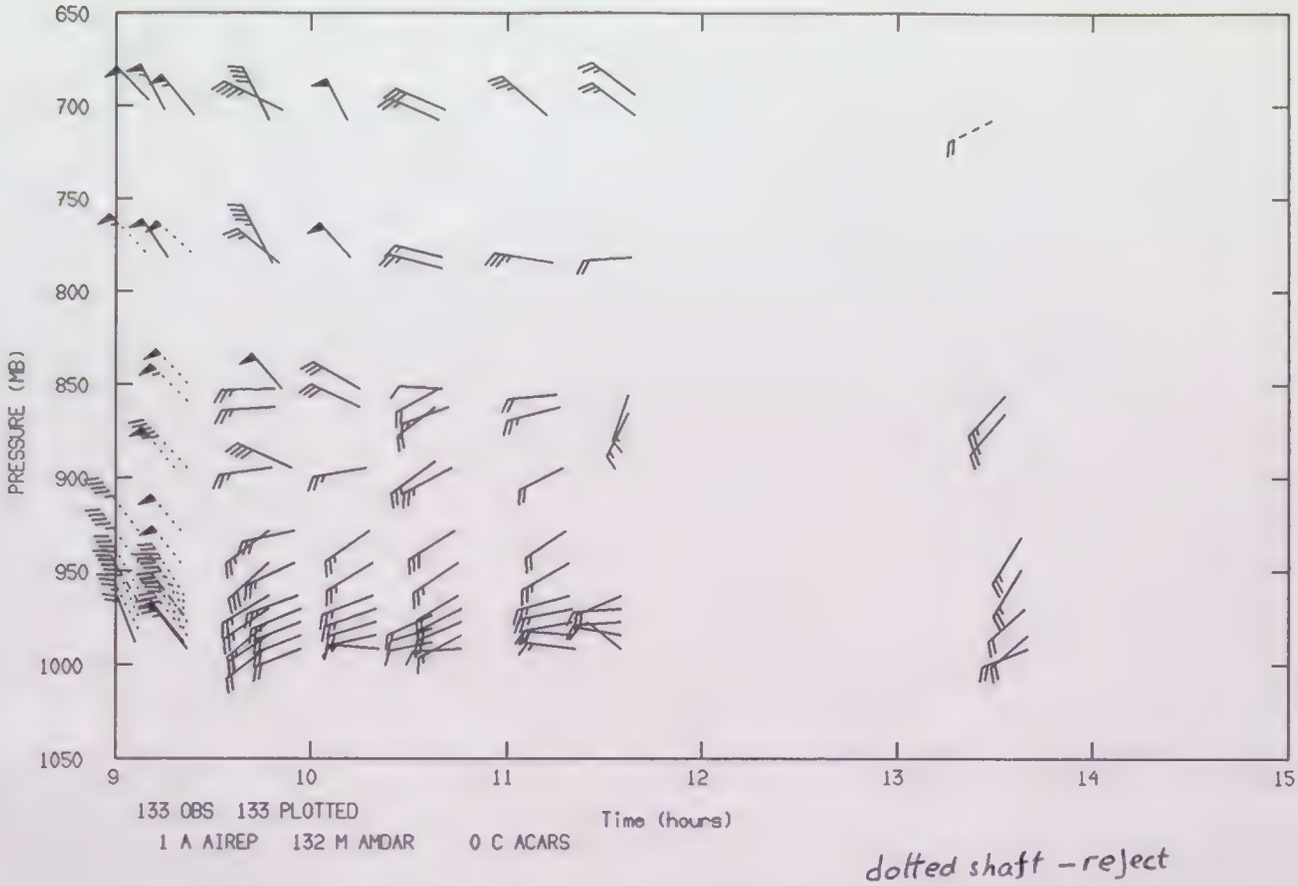


19 OBS 19 PLOTTED
19 A AIREP 0 C ACARS

dotted shaft - reject
dashed shaft - suspect

Figure 4. Melbourne airport case.

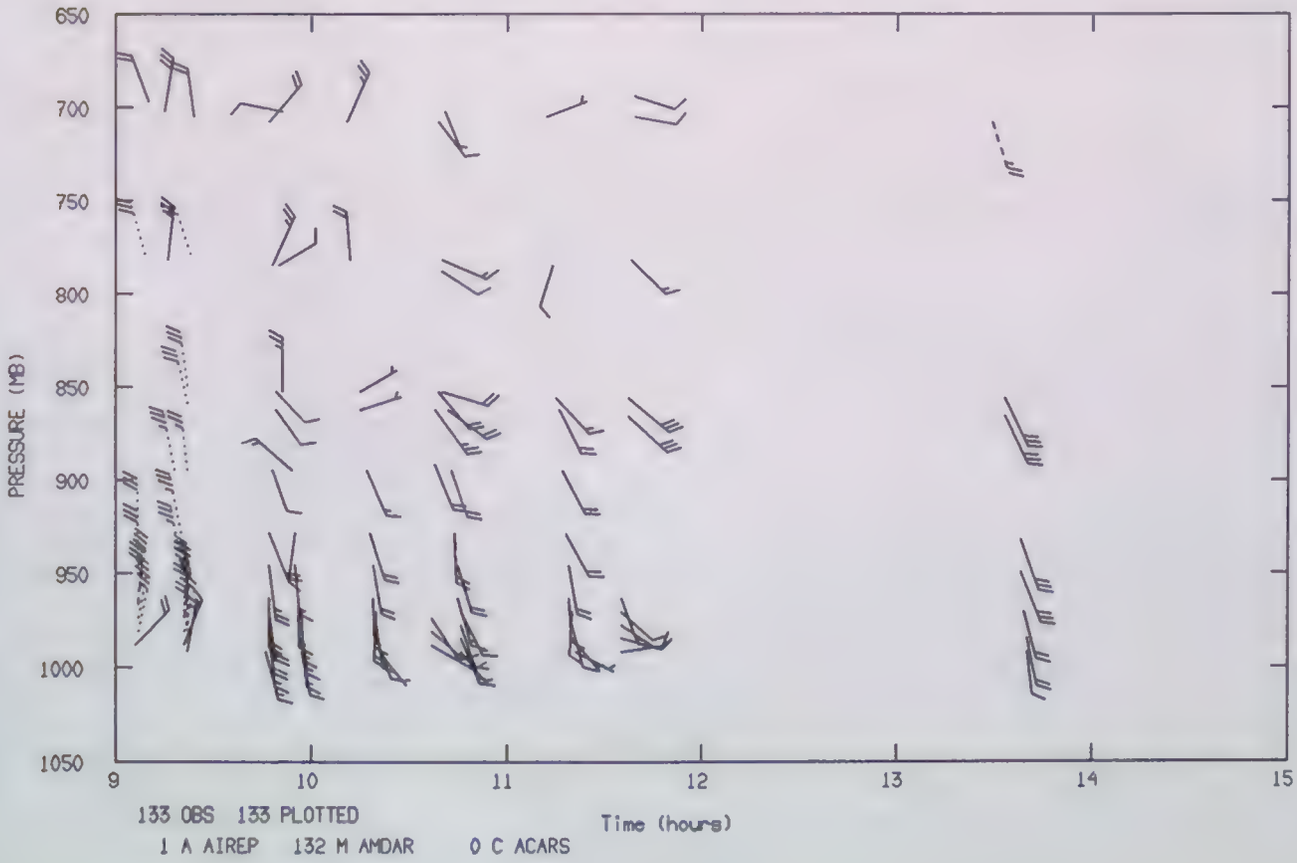
a) OPD OBSERVATIONS 1050.0 MB TO 650.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE -37.7 144.8 AREA: 1.10



133 OBS 133 PLOTTED
1 A AIREP 132 M AMDAR 0 C ACARS

dotted shaft - reject
dashed shaft - suspect

b) OPD (OBS. - BK.) 1050.0 MB TO 650.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE -37.7 144.8 AREA: 1.10



133 OBS 133 PLOTTED
1 A AIREP 132 M AMDAR 0 C ACARS

OPD OBSERVATIONS 1050.0 MB TO 650.0 MB. SAMPLE

WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00

12Z 28 / 9 / 1992. GRID CENTRE -40.0 145.0 SCALE: 4.00

850 MB. T + 6 PPFILE1 HEIGHT

Figure 5. Melborne airport case.
All the rejected and suspect observations are shown, but only a sample of other data.

CONTOUR INTERVAL: 60 M

654 OBS 116 PLOTTED

7 R TEMP 15 P PILOT 0 U TSHIP 0 A AIREP 27 M AMDAR

0 C ACARS 67 S SATOB

dotted shaft - reject
dashedshaft - suspect

OPD OBSERVATIONS 350.0 MB TO 150.0 MB. ALL

WIND (FULL FEATHER = 10 KTS) 28SEP92:12.00 - 28SEP92:15.00

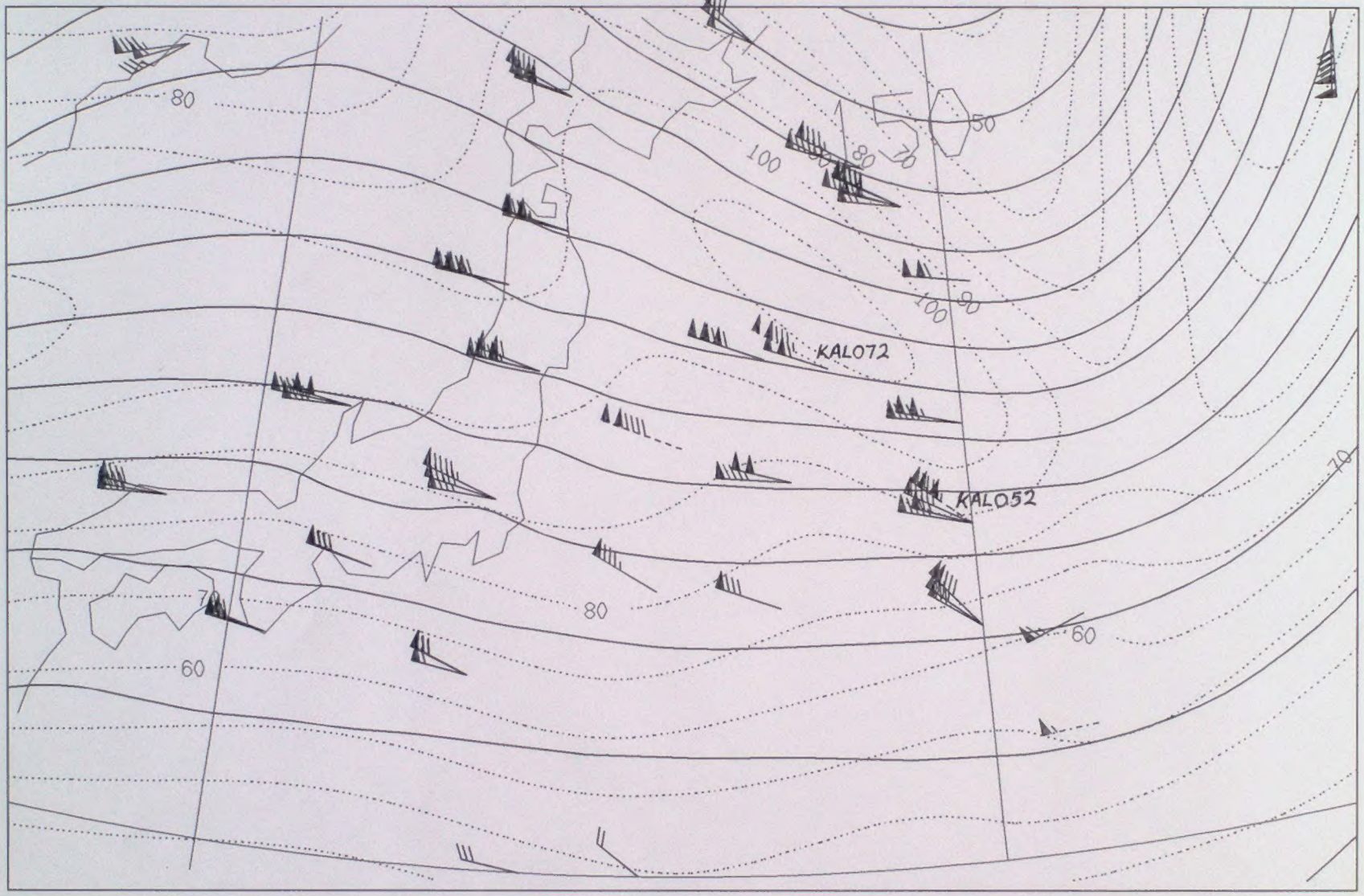
12Z 28 / 9 / 1992. GRID CENTRE 37.0 144.0 SCALE: 8.00

Figure 6. KAL072 and KAL052 and neighbours.

(note observations before 12Z not plotted)

250 MB. T + 6 PPFILE1 HEIGHT

T + 6 PPFILE1 WIND SPEED



CONTOUR INTERVAL: 60 M

CONTOUR INTERVAL: 10 KNOTS

76 OBS 76 PLOTTED

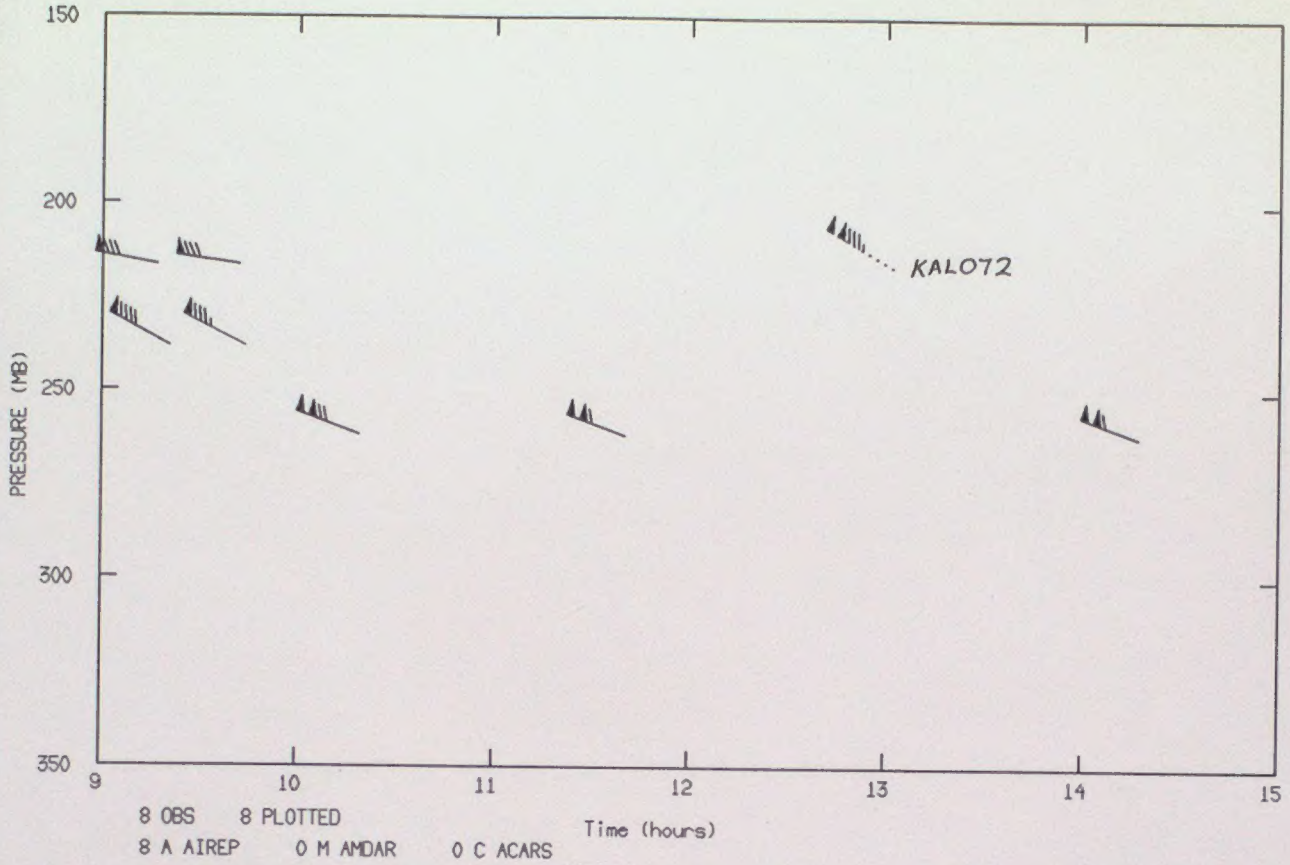
39 R TEMP 0 P PILOT 0 U TSHIP 36 A AIREP 0 M AMDAR

0 C ACARS 1 S SATOB

dotted shaft - rejected, dashed - suspect, dash-dotted - prior reject (blacklisted)

OPD OBSERVATIONS 350.0 MB TO 150.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE 38.3 147.2 AREA: 1.10

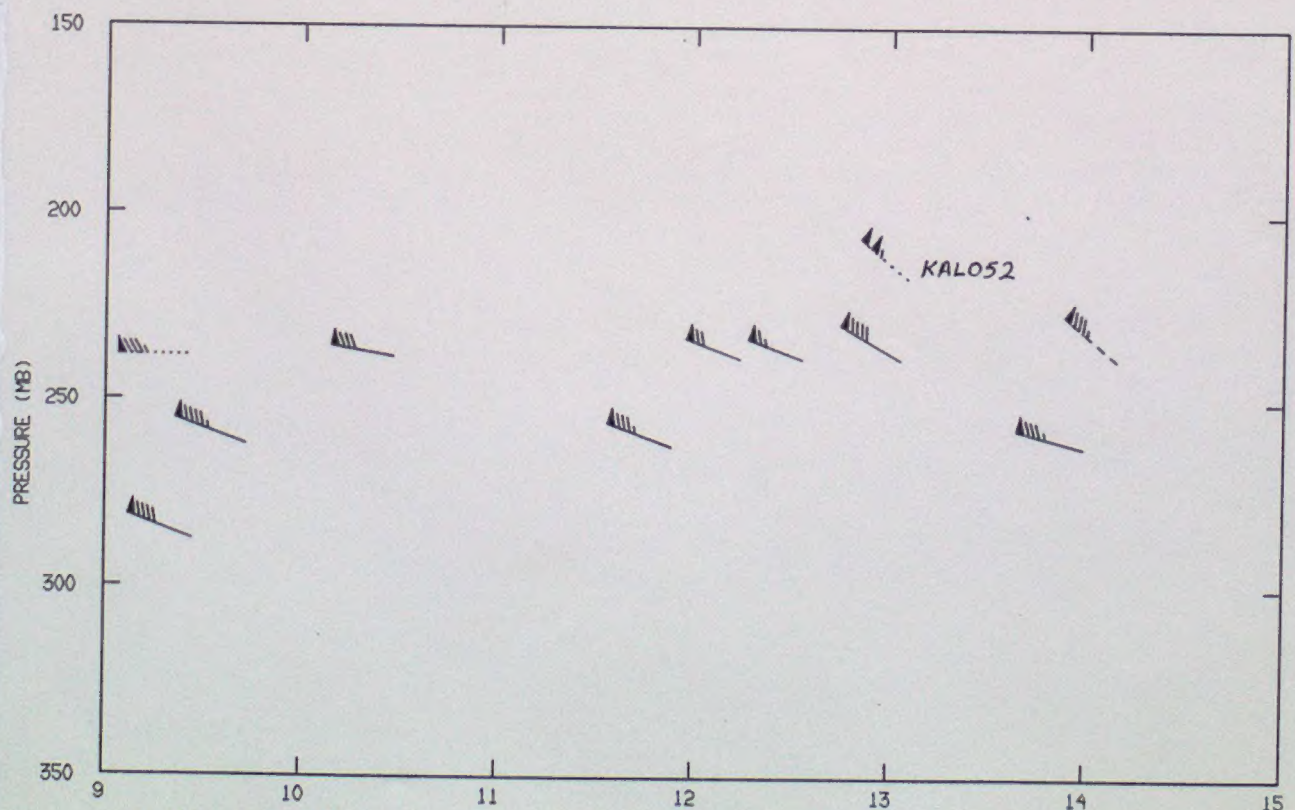
Figure 7. KAL072 and neighbours.



8 OBS 8 PLOTTED
8 A AIREP 0 M AMDAR 0 C ACARS

OPD OBSERVATIONS 350.0 MB TO 150.0 MB. ALL
WIND (FULL FEATHER = 10 KTS) 28SEP92:09.00 - 28SEP92:15.00
12Z 28 / 9 / 1992. GRID CENTRE 35.5 150.0 AREA: 1.10

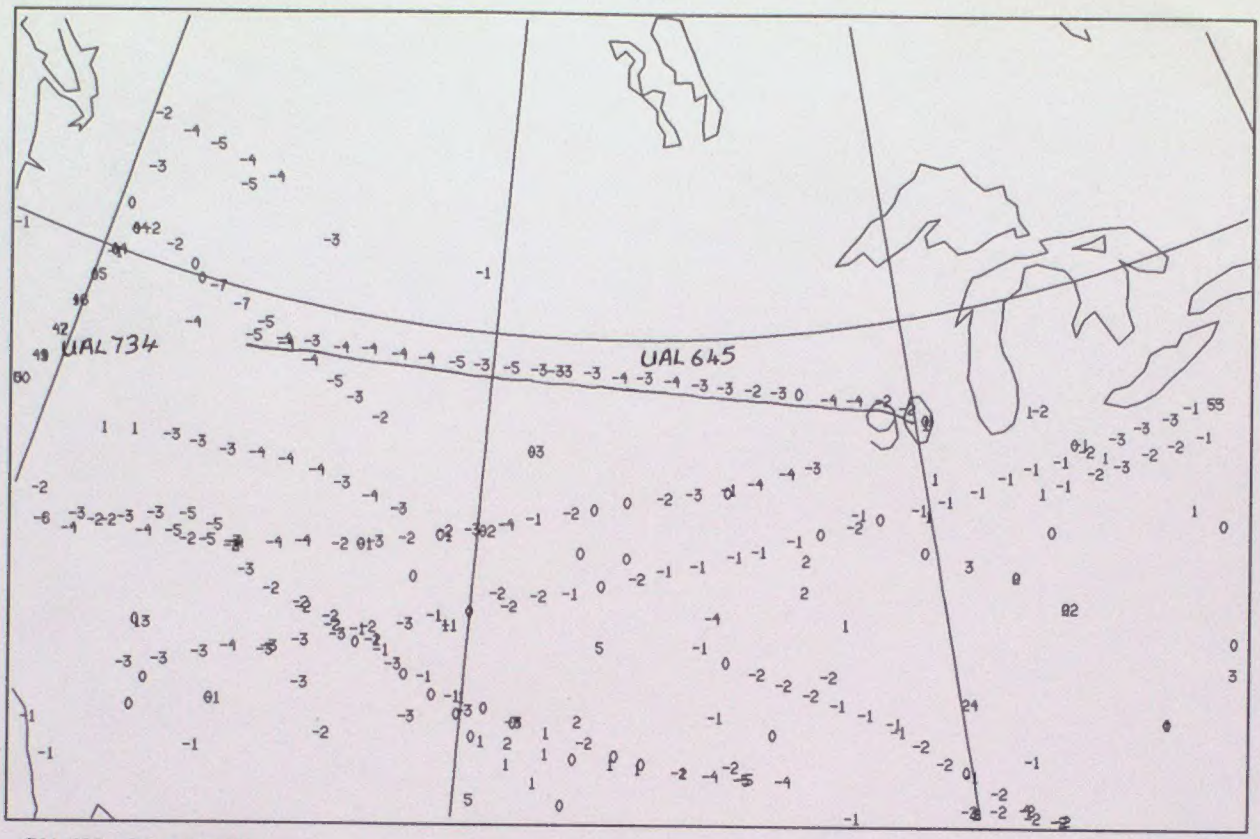
Figure 8. KAL052 and neigbours.



11 OBS 11 PLOTTED
11 A AIREP 0 M AMDAR 0 C ACARS

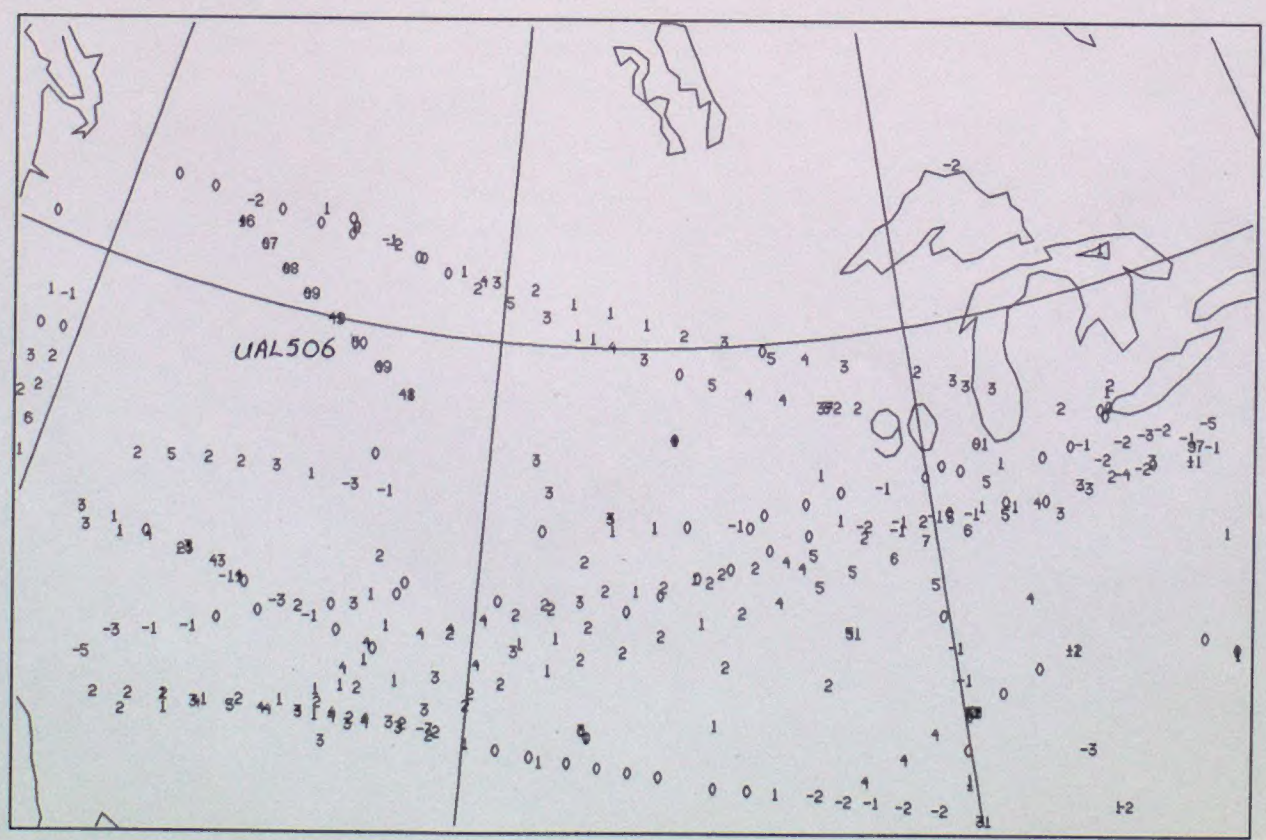
dotted shaft - reject
dashed shaft - suspect

a) OPD (OBS. - BK.) 275.0 MB TO 225.0 MB. ALL *Figure 9. Temperature bias case.*
TEMPERATURE 28JAN93:15.00 - 28JAN93:21.00 *(note that only one temperature from UAL645 was*
GRID CENTRE 43.0 -100.0 SCALE: 6.00 *rejected operationally).*



301 OBS 301 PLOTTED
0 R TEMP 0 P PILOT 301 A AIREP 0 C ACARS

b) OPD (OBS. - BK.) 225.0 MB TO 175.0 MB. ALL
TEMPERATURE 28JAN93:15.00 - 28JAN93:21.00
GRID CENTRE 43.0 -100.0 SCALE: 6.00



341 OBS 341 PLOTTED
0 R TEMP 0 P PILOT 341 A AIREP 0 C ACARS